Sample Reuse Selection and Allocation Criteria

by

Seymour Geisser*
University of Minnesota

Technical Report No. 325
August 1978

Sample Reuse Selection and Allocation Criteria

by

Seymour Geisser*
University of Minnesota

1. Introduction

We present several sample reuse criteria for selecting among alternative densities, or probability functions, those most appropriate for classifying or allocating new observations. In standard model selection problems, when densities are completely specified, the choice between models can appropriately be made to rest on a comparison of the alternative probability densities of the sample as determined by the models. Assume that there are two distinct populations π_1 and π_2 from which sets of values have been observed $X_1 = x_1$ and $X_2 = x_2$ respectively with designation for the joint set of random variables $X = (X_1, X_2)$ and $X_i = \{X_{i1}, \ldots, X_{iN_i}\}$, $i = 1,2$. Then we need to determine the most appropriate density $f_\omega(x|\pi_i)$, indexed by the double designator ω, $\omega \in \Omega$, which jointly specifies a pair of densities for π_1 and π_2, where Ω represents the totality of such pairs of potential predicting densities under consideration. This is accomplished by obtaining that ω which maximizes

$$p(\omega) f_\omega(x_1, x_2|\pi_1,\pi_2) \tag{1.1}$$

w.r.t. ω for $p(\omega)$ the prior probability that ω is the correct density pair.

Of course, if we know the true ω^*, then the allocation or diagnosis of a new observation $Z = z$ should depend on the posterior odds ratio

*This work was supported in part by a NIH-GMS research grant.

$$R(\omega^*,z) = \frac{q_1 f_{\omega^*}(z|\pi_1)}{q_2 f_{\omega^*}(z|\pi_2)} = \frac{Pr[\pi_1|z,\omega^*]}{Pr[\pi_2|z,\omega^*]} \tag{1.2}$$

where $q_i = Pr[\pi_i]$, the prior probability that $z \in \pi_i$.

More generally, we can incorporate available probabilistic information on ω. Assuming q_i is independent of ω and noting that

$$p(\omega|z) = f(z|\omega)p(\omega)/f(\omega),$$

then we can calculate

$$Pr[\pi_i|z] = E_{\omega|z}[Pr(\pi_i|z,\omega)] \propto q_i E_{\omega}[f_{\omega}(z|\pi_i)]$$

and

$$R(z) = Pr[\pi_i|z]/Pr[\pi_i|z] ,$$

where the first expectation is over the conditional probability of ω given z and the second is merely over the marginal probability function $p(\omega)$. In practice ω would usually range over a small finite number of discrete possibilities so that

$$Pr[\pi_i|z] \propto q_i \sum_{k=1}^{K} f_{\omega_k}(z|\pi_i)p(\omega_k) \tag{1.3}$$

would be a simple mixture of densities. Although , assuming we know $p(\omega)$, there is no intrinsic difficulty in applying (1.3) directly for classification purposes, our goal is often a compromise in that we would both like to select a model and then use that model for classification. When this is the case we shall not average over ω , as it were, but make the choice of ω in keeping with our ultimate goal of classification.

To further complicate the issue it is a fact that in most instances a designator ω is a specification of the form of the density with values of the parameters unknown. Such cases then would require that prior distributions for the parameters be introduced and predictive (marginal) densities be calculated for X for each ω . Maximization of

$$p(\omega)f_{\omega}(x_1,x_2|\pi_1,\pi_2) \tag{1.4}$$

w.r.t. ω, where, assuming the X_{ij}'s are independently distributed,

$$f_\omega(x_1,x_2|\pi_1,\pi_2) = \int \prod_{i=1}^{2} \prod_{j=1}^{N_i} f_\omega(x_{ij}|\theta_i,\pi_i) g_\omega(\theta) d\theta \qquad (1.5)$$

and $g_\omega(\theta)$ is the prior density of the set of parameters specified by ω, π_1 and π_2 and $\theta = \theta_1 \cup \theta_2$ represents the set of distinct parameters, leads to a full Bayesian solution for the selection problem. But this solution usually requires proper prior distributions with hyperparameters specified.

As before, if a choice of model $\omega = \omega^*$ is made and one uses this for allocation, the relevant posterior odds used for assigning a new observation z, are

$$R(\omega^*,z,x) = \frac{q_1 f_{\omega^*}(z|x,\pi_1)}{q_2 f_{\omega^*}(z|x,\pi_2)} \quad , \qquad (1.6)$$

where

$$f_\omega(z|x,\pi_i) = \int f_\omega(z|\theta_i,\pi_i) dG_\omega(\theta|x) \qquad (1.7)$$

and

$$dG_\omega(\theta|x) \propto g_\omega(\theta) \prod_i \prod_j f_\omega(x_{ij}|\theta_i,\pi_i) d\theta \, . \qquad (1.8)$$

If the complete Bayesian solution is to be used for classification in the presence of a prior probability function for ω, then one needs to calculate

$$\Pr[\pi_i|z,x,q_i] \propto q_i E_\omega f_\omega(z|x,\pi_i) f_\omega(x_1,x_2|\pi_1,\pi_2) \qquad (1.9)$$

as the basis for assigning z, where the expectation is over the prior distribution of ω. One can justify (1.9) in the following way: Since

$$\Pr[\pi_i|z,x,q_i,\omega] = \frac{q_i f_\omega(z|x,\pi_i)}{f_\omega(z|x)} \qquad (1.10)$$

where

$$f_\omega(z|x) = q_1 f_\omega(z|x,\pi_1) + q_2 f_\omega(z|x,\pi_2) ,$$

then clearly

$$\Pr[\pi_i|z,x,q_i] \propto q_i E_{\omega|z,x}\left(\frac{f_\omega(z|x,\pi_i)}{f_\omega(z|x)}\right) \tag{1.11}$$

i.e. the expectation on the r.h.s. is w.r.t. the distribution of ω conditional on $Z = z$ and $X = x$. Now

$$p(\omega|z,x) = \frac{f_\omega(z|x) f_\omega(x) p(\omega)}{f(z|x) f(x)} \tag{1.12}$$

$$\left.\begin{aligned} f_\omega(x) &= \int \prod_i \prod_j f_\omega(x_{ij}|\theta_i,\pi_i) dG_\omega(\theta) = f_\omega(x_1,x_2|\pi_1,\pi_2) \\ p(\omega|x) &= \frac{f_\omega(x) p(\omega)}{f(x)} = \frac{f_\omega(x_1,x_2|\pi_1,\pi_2) p(\omega)}{f(x)} \\ f(x) &= \int f_\omega(x_1,x_2|\pi_1,\pi_2) dP(\omega) . \end{aligned}\right\} \tag{1.13}$$

After evaluating (1.12), using the results of (1.13), we evaluate (1.11) and obtain (1.9).

Full Bayesian solutions require a body of prior knowledge that is often unavailable. Also, even when presumed available, the analysis may be highly sensitive to some of the assumptions which, in fact, may have been grossly violated.

For the aforementioned problem, data analytic solutions based on sample reuse techniques, in the spirit of Geisser and Eddy (1977), will be presented. We consider then the case of two distinct populations that have been sampled with respect to some set of p variables but there is some doubt as to the distributions which generated the samples.

Hence, to the usual classification problem there is added the uncertainty of distributional assumptions regarding the two populations and a goal of the model selection procedure is to optimize classification in some sense--i.e. select the single model which will do the best job of classifying future observations.

2. Criteria for Model Selection

Previously, Geisser and Eddy (1977) recommended for model selection geared to prediction, that a series of predicting densities $f_\omega(z|x,\pi_i)$ be established from which to compute $R(\omega^*,z)$ of (1.6). One then maximizes the reused product of conditional predicting densities of this form to obtain ω^*. In this case, this would be equivalent to maximizing w.r.t. ω, $p(\omega)L(\omega)$ where

$$L_\omega = \prod_{i=1}^{2} \prod_{j=1}^{N_i} f_\omega(x_{ij}|x_{(ij)},\pi_i) \tag{2.1}$$

and $X_{(ij)} = x_{(ij)}$ the set of observations $X = x$ with $X_{ij} = x_{ij}$ deleted but of the same form as $f_\omega(z|x,\pi_i)$. This is certainly a useful procedure for rather tight specifications when they are met. A variant of this procedure which tends more to emphasize the ultimate goal, classification of a new observation, is maximization of $p(\omega)O(\omega)$ where

$$O(\omega) = \frac{L(\omega)}{\bar{L}(\omega)} \left(\frac{q_1}{q_2}\right)^{N_1-N_2} \tag{2.2}$$

and

$$\bar{L}_\omega = \prod_{i=1}^{2} \prod_{j=1}^{N_i} f_\omega(x_{ij}|x_{(ij)},\pi_{3-i}) \ . \tag{2.3}$$

This reflects more directly the posterior odds ratio for a given ω which will be used to classify a new observation. Note that $(q_1/q_2)^{N_1-N_2}$ is independent of ω and therefore does not effect comparisons for various ω .

Both (2.1) and (2.2) depend on products--perhaps too much so to be usefully robust to the presence of outliers, contaminants or inadvertantly misclassified observations in the initial or training samples. This has

the effect that a single observation in a low density region has enormous influence on $L(\omega)$. Hence even if a particular ω^* were "true," the insinuation of a single wildly discrepant observation could so diminish $L(\omega^*)$ that other ω could wrongly come to the fore. The effect of such an observation is mitigated by the use of $O(\omega)$ if it is in a low density region for both specifications of ω^*, since the odds ratio for the discrepant observation would minimally influence $O(\omega^*)$. Hence the effect of a few discrepant observations of this type would be largely diluted. On the other hand observations that were actually misclassified in a high density region of one of the pair and in a low density region of the other would have enormous effect on $O(\omega)$ in the wrong direction--even more so than on L_ω . Hence both $L(\omega)$ and $O(\omega)$ are highly volatile criteria in that they tend to be somewhat oversensitive to aberrancies.

Because of this we present a third sample reuse criterion which is far less sensitive to the type of eccentricities previously discussed--in that an aberrant observation would have far less influence on the resolution of the appropriate ω .

Let

$$R_{ij}(\omega) = \frac{q_1 f_\omega(x_{ij}|x_{(ij)},\pi_1)}{q_2 f_\omega(x_{ij}|x_{(ij)},\pi_2)} \tag{2.4}$$

for $j = 1, \ldots, N_i$; $i = 1,2$. Thus x_{ij} would be assigned to π_1 or π_2 as $R_{ij}(\omega) \geq 1$ or < 1 . Hence for each ω , $R_{ij}(\omega)$ will correctly assign $n_i(\omega)$ of the x_{ij}'s to π_i . Optimization then requires that we choose that ω which maximizes $Q(\omega) = q_1 N_1^{-1} n_1(\omega) + q_2 N_2^{-1} n_2(\omega)$, or $Q(\omega)$ may be multiplied by $p(\omega)$, if a sensible $p(\omega)$ exists for the

alternative ω, before maximizing. If q_i is unknown but estimable by $N_i/(N_1 + N_2)$, then for constant $p(\omega)$, we are maximizing $n_1(\omega)+n_2(\omega)$, the total number correctly classified by designator ω. Note also that if there is an ω such that $n_i(\omega^*) \geq n_i(\omega)$, $i = 1,2$, then irrespective of q_i and N_i this ω^* is optimal.

Although with probability 1, unique solutions for ω's specified by continuous densities can be guaranteed for (2.1) and (2.2), this is obviously not the case for (2.4). Hence if the maximum is achieved for several ω's using (2.4) these "ties" may be broken by use of (2.1) or (2.2) to discriminate amongst these ω's. Further even when unique solutions exist in regard to any of the selection criterion they may not be distinguishable for classification purposes. For example, assume $N_1 = N_2 = M$ and either $q_1 = q_2$ or $\hat{q}_1 = \hat{q}_2$ and the possible models are: 1) a pair of normal densities with differing unknown means but with the same but unknown variance that is estimated by insertion of the usual estimators in the normal densities,

$$f_{\omega_\varphi}(z|\pi_i) = \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{1}{2s^2}(z-\bar{x}_i)^2} \qquad i = 1,2 \qquad (2.5)$$

where

$$s^2 = (2M-2)^{-1} \sum_{i=1}^{2} \sum_{j=1}^{M} (x_{ij} - \bar{x}_i)^2, \text{ and } \bar{x}_i = M^{-1} \sum_{i=1}^{M} x_i ;$$

or 2) a pair of t densities (which can also be considered Bayesian estimates of the underlying normal pair, Geisser (1971), Aitchison (1975)),

$$f_{\omega_t}(z|\pi_i) = \left[\frac{M}{\pi(2M-2)(M+1)}\right]^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(2M-1))}{s\Gamma(\frac{1}{2}(2M-2))} \left[1 + \frac{M(z-\bar{x}_i)^2}{(M+1)(2M-2)s^2}\right]^{-\frac{1}{2}(2M-1)} . \qquad (2.6)$$

It is then easy to demonstrate that for every z, $R(z,\omega_\varphi)$ results in the same allocation as $R(z,\omega_t)$ since

$$R(z,\omega_\varphi) = \frac{f_{\omega_\varphi}(z|\pi_1)}{f_{\omega_\varphi}(z|\pi_2)} > 1 \quad \text{or} \quad \leq 1$$

implies that

$$R(z,\omega_t) = \frac{f_{\omega_t}(z|\pi_1)}{f_{\omega_t}(z|\pi_2)} > 1 \quad \text{or} \quad \leq 1 \quad \text{respectively.}$$

It is to be noted, however, that the value of the odds ratio itself varies for the two alternative forms so that a different cutoff point would result in differing allocations. This would occur if either for known q_1 and q_2, $q_1 \neq q_2$ or if unknown and estimated $\hat{q}_1 \neq \hat{q}_2$.

3. An Application to Multivariate Normal Populations

Assume that under ω_1, $\underset{\sim}{x}_1 = \{\underset{\sim}{x}_{11}, \ldots, \underset{\sim}{x}_{1N_1}\}$ and $\underset{\sim}{x}_2 = \{\underset{\sim}{x}_{21}, \ldots, \underset{\sim}{x}_{2N}\}$ are respectively the observed values of independently distributed p-dimensional random variables which, respectively under π_1, arose from a $N(\underset{\sim}{\mu}_1,\Sigma)$, and under π_2, arose from a $N(\underset{\sim}{\mu}_2,\Sigma)$. Under ω_2 similarly the set of observations $\underset{\sim}{x}_1$ and $\underset{\sim}{x}_2$ are the observed sets of values which respectively arose as independent realizations of a $N(\underset{\sim}{\mu}_1,\Sigma_1)$ under π_1 and a $N(\underset{\sim}{\mu}_2,\Sigma_2)$ under π_2.

For the classification of a future vector observation $\underset{\sim}{Z} = \underset{\sim}{z}$, where $\Pr(\pi_1) = q_1$, and $q_2 = \Pr(\pi_2)$, $q_1 + q_2 = 1$ and convenient prior density; under ω_1,

$$g_{\omega_1}(\underset{\sim}{\mu}_1,\underset{\sim}{\mu}_2,\Sigma^{-1}) \propto |\Sigma|^{(p+1)/2},$$

Geisser (1964) obtained

$$\Pr(\pi_i|\underset{\sim}{z}, \omega_1) \quad \propto \quad q_i f_{\omega_1}(\underset{\sim}{z}|\bar{\underset{\sim}{x}}_i, S, N_i, N, \pi_i)$$

where

$$f_{\omega_1}(\underset{\sim}{z}|\bar{\underset{\sim}{x}}_i, S, N_i, N, \pi_i) = \left[\frac{N_i}{\pi(N_i-1)}\right]^{p/2} \frac{\Gamma(\frac{N-1}{2})}{\Gamma(\frac{N-p-1}{2})|(N-2)S|^{\frac{1}{2}}} \left[1 + \frac{N_i(\underset{\sim}{z}-\bar{\underset{\sim}{x}}_i)'S^{-1}(\underset{\sim}{z}-\underset{\sim}{x}_i)}{(N_i+1)(N-2)}\right]^{-(N-1)/2} \tag{3.1}$$

and

$$\bar{\underset{\sim}{x}}_i = N_i^{-1} \sum_{j=1}^{N_i} \underset{\sim}{x}_{ij}, \quad (N_i-1)\, S_i = \sum_{j=1}^{N_i} (\underset{\sim}{x}_{ij} - \bar{\underset{\sim}{x}}_i)(\underset{\sim}{x}_{ij} - \bar{\underset{\sim}{x}}_i)',$$

$$(N-2)S = (N_1-1)S_1 + (N_2-1)S_2, \quad N = N_1 + N_2 \ .$$

This predictive density (3.1), suggested by Geisser (1964, 1971) as a Bayesian estimate of the sampling density, was actually shown by Murray (1977) to be optimal in the frequency sense among all estimators of the sample density that are invariant under translations and non-singular linear transformations of the sample space using as a goodness of fit criterion the information measure of Kullback and Liebler (1951).

Under ω_2 we assume a similar prior density

$$g_{\omega_2}(\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \Sigma_1, \Sigma_2) \quad \propto \quad |\Sigma_1|^{(p+1)/2}|\Sigma_2|^{(p+1)/2}$$

and obtain

$$\Pr(\pi_i|\underset{\sim}{z}, \omega_2) \quad \propto q_i f_{\omega_2}(\underset{\sim}{z}|\bar{\underset{\sim}{x}}_i, S_i, N_i, \pi_i)$$

where

$$f_{\omega_2}(\underset{\sim}{z}|\bar{\underset{\sim}{x}}_i, S_i, N_i, \pi_i) = \left(\frac{N_i}{\pi(N_i+1)}\right)^{\frac{p}{2}} \frac{\Gamma(\frac{N_i}{2})}{\Gamma(\frac{N_i-p}{2})|(N_i-1)S_i|^{\frac{1}{2}}} \left[1 + \frac{N_i(\underset{\sim}{z}-\bar{\underset{\sim}{x}}_i)'S_i^{-1}(\underset{\sim}{z}-\bar{\underset{\sim}{x}}_i)}{(N_i+1)(N_i-1)}\right]^{-N_i/2} . \tag{3.2}$$

To establish which of the two models, ω_1 or ω_2, is more appropriate for classification, we can apply one or more of the following methods.

Method I. Let

$$P(\omega_k) = p(\omega_k)L(\omega_k) \qquad k = 1,2$$

where

$$\begin{aligned} L(\omega_1) &= \prod_{i=1}^{2}\prod_{j=1}^{N_i} f_{\omega_1}(\underset{\sim}{x}_{ij}|\bar{\underset{\sim}{x}}_{i(j)}, S_{(ij)}, N_i-1, N-1, \pi_i) \\ L(\omega_2) &= \prod_{i=1}^{2}\prod_{j=1}^{N_i} f_{\omega_2}(\underset{\sim}{x}_{ij}|\bar{\underset{\sim}{x}}_{i(j)}, S_{i(j)}, N_i-1, \pi_i) \end{aligned} \tag{3.3}$$

where $f_{\omega_1}(\cdot)$ and $f_{\omega_2}(\cdot)$ are defined as in (3.1) and (3.2) respectively and

$$\bar{\underset{\sim}{x}}_{i(j)} = (N_i-1)^{-1}[N\bar{\underset{\sim}{x}}_i - \underset{\sim}{x}_{ij}]$$

$$(N-3)S_{(ij)} = (N_i-2)S_{i(j)} + (N_{3-i}-1)S_{3-i} \qquad i = 1,2$$

$$(N_i-2)S_{i(j)} = \sum_{\substack{t=1 \\ t\neq j}}^{N_i} (\underset{\sim}{x}_{it}-\bar{\underset{\sim}{x}}_{i(j)})(\underset{\sim}{x}_{it}-\bar{\underset{\sim}{x}}_{i(j)})' .$$

Select that ω_k which maximizes $P(\omega_k)$.

Method II.

Define

$$\begin{aligned} \bar{L}(\omega_1) &= \prod_{i=1}^{2}\prod_{j=1}^{N_i} f_{\omega_1}(\underset{\sim}{x}_{ij}|\bar{\underset{\sim}{x}}_{3-i}, S_{(ij)}, N_{3-i}, N-1, \pi_i) \\ \bar{L}(\omega_2) &= \prod_{i=1}^{2}\prod_{j=1}^{N_i} f_{\omega_2}(\underset{\sim}{x}_{ij}|\bar{\underset{\sim}{x}}_{3-i}, S_{3-i}, N_{3-i}, \pi_i) \end{aligned} \tag{3.4}$$

and

$$O(\omega_k) = \frac{L(\omega_k)}{\bar{L}(\omega_k)} \cdot \left(\frac{q_1}{q_2}\right)^{N_1-N_2} \qquad k = 1,2 \tag{3.5}$$

select that ω_k which maximizes $p(\omega_k)O(\omega_k)$

Method III.

Define

$$\left.\begin{aligned} R_{1j}(\omega_1) &= \frac{q_1 f_{\omega_1}(\underset{\sim}{x}_{1j}|\bar{\underset{\sim}{x}}_{1(j)}, S_{(1j)}, N_1-1, N-1, \pi_1)}{q_2 f_{\omega_1}(\underset{\sim}{x}_{1j}|\bar{\underset{\sim}{x}}_{2}, S_{(1j)}, N_2, N-1, \pi_2)} \\ R_{2j}(\omega_1) &= \frac{q_1 f_{\omega_1}(\underset{\sim}{x}_{2j}|\bar{\underset{\sim}{x}}_{1}, S_{(2j)}, N_1, N-1, \pi_1)}{q_2 f_{\omega_1}(\underset{\sim}{x}_{2j}|\bar{\underset{\sim}{x}}_{2(j)}, S_{(2j)}, N_2-1, N-1, \pi_2)} \end{aligned}\right\} \quad (3.6)$$

$$\left.\begin{aligned} R_{1j}(\omega_2) &= \frac{q_1 f_{\omega_2}(\underset{\sim}{x}_{1j}|\bar{\underset{\sim}{x}}_{1(j)}, S_{1(j)}, N_1-1, \pi_1)}{q_2 f_{\omega_2}(\underset{\sim}{x}_{ij}|\bar{\underset{\sim}{x}}_{2}, S_2, N_2, \pi_2)} \\ R_{2j}(\omega_2) &= \frac{q_1 f_{\omega_2}(\underset{\sim}{x}_{2j}|\bar{\underset{\sim}{x}}_{1}, S_1, N_1, \pi_1)}{q_2 f_{\omega_2}(\underset{\sim}{x}_{2j}|\bar{\underset{\sim}{x}}_{2(j)}, S_{2(j)}, N_2-1, \pi_2)} \end{aligned}\right\} \quad (3.7)$$

Now calculate

$$Q(\omega_k) = q_1 n_1(\omega_k) N_1^{-1} + q_2 n_2(\omega_k) N_2^{-1}$$

where, in general,

$$n_i(\omega_k) = \sum_{j=1}^{N_i} \delta_{ij}(\omega_k) \quad (3.8)$$

and

$$\delta_{ij}(\omega_k) = \begin{cases} 1 & \text{if } [R_{ij}(\omega_k)]^{3-2i} > 1 \\ 0 & \text{otherwise} \end{cases} \quad (3.9)$$

and incidentally

$$O(\omega_k) = \prod_{i=1}^{2} \prod_{j=1}^{N_i} [R_{ij}(\omega_k)]^{3-2i} \quad (3.10)$$

Further choose that ω_k which maximizes $p(\omega_k)Q(\omega_k)$, or if $\hat{q}_i = N_i/N$ and one is a priori indifferent to a choice between ω_1 and ω_2, select that ω_k which maximizes

$$n(\omega_k) = n_1(\omega_k) + n_2(\omega_k) \quad (3.11)$$

the total number of $\underset{\sim}{x}_{ij}$'s correctly classified using $R_{ij}(\omega_k)$.

4. Summary

The three sample reuse methods presented here approximately simulate different comparative measures. The first approximates the posterior probability assuming the source, π_i, of the omitted observation is known, which heavily emphasizes selecting the best model. The second attempts to approximate the posterior odds ratio that each observation is correctly classified and uses as a measure the product of these odds ratios.

The last initially treats the omitted observation as if its source were unknown and proceeds to simulate the classification scheme itself by assigning each omitted observation to π_1 or π_2 and then determines the number correctly classified for each ω_k. This permits each observation to contribute more equally to the selection measure.

Actually for any particular problem, given the kind of computations to be made, it would appear that all three methods can be simultaneously calculated and their results compared before a final conclusion is reached as to the choice of ω most suitable for allocating new observations.

References

Aitchison, J. (1975). Goodness of prediction fit. Biometrika, 62, 547-54.

Geisser, S. (1964). Posterior odds for multivariate normal classifications, J. Roy. Statist. Soc. B, 25, 368-376.

Geisser, S. (1971). The inferential use of predictive distributions. Foundations of Statistical Inference edited by V. Godambe and D. Sprott. Holt, Rinehart and Winston, 456-469.

Geisser, S. and Eddy, W. F. (1978). A predictive approach to model selection. J. Amer. Statist. Assoc. (in press).

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. Ann. Math. Statist., 22, 79-86.

Murray, G. D. (1977). A note on the estimation of probability density functions. Biometrika, 64, 1, 150-2.